

ДИФРАКЦИЯ И РАССЕЯНИЕ ИОНИЗИРУЮЩИХ ИЗЛУЧЕНИЙ

УДК 539.26

ВОССТАНОВЛЕНИЕ ПРОФИЛЕЙ ИНТЕНСИВНОСТИ МАЛОУГЛОВОГО РАССЕЯНИЯ ОТ ДВУХ КОНФОРМАЦИОННЫХ СОСТОЯНИЙ 3-ИЗОПРОПИЛМАЛАТДЕГИДРОГЕНАЗЫ С ПОМОЩЬЮ ЭВОЛЮЦИОННОГО ФАКТОРНОГО АНАЛИЗА

© 2023 г. П. В. Конарев^{1,2,*}, В. В. Волков^{1,2}

¹Институт кристаллографии им. А.В. Шубникова, ФНИЦ “Кристаллография и фотоника” РАН, Москва, Россия

²Национальный исследовательский центр “Курчатовский институт”, Москва, Россия

*E-mail: peter_konarev@mail.ru

Поступила в редакцию 10.05.2023 г.

После доработки 25.05.2023 г.

Принята к публикации 25.05.2023 г.

Проведено моделирование разделения перекрывающихся вкладов в интенсивность малоуглового рассеяния от молекул 3-изопропилмалатдегидрогеназы из *Thermus thermophilus*, находящихся в растворе в двух конформационных состояниях, при использовании хроматографической колонки, присоединенной к измерительной кювете синхротронной станции малоуглового рентгеновского рассеяния. По теоретическим наборам малоугловых данных, к которым был добавлен пуассоновский шум в пределах 3–5%, восстановлены профили интенсивности рассеяния от открытой и закрытой форм молекул фермента методом эволюционного факторного анализа. Таким образом, подтверждена эффективность работы данного алгоритма при разделении вкладов компонентов в случае смесей, состоящих из частиц с одинаковой молекулярной массой.

DOI: 10.31857/S0023476123600155, EDN: XRTPYF

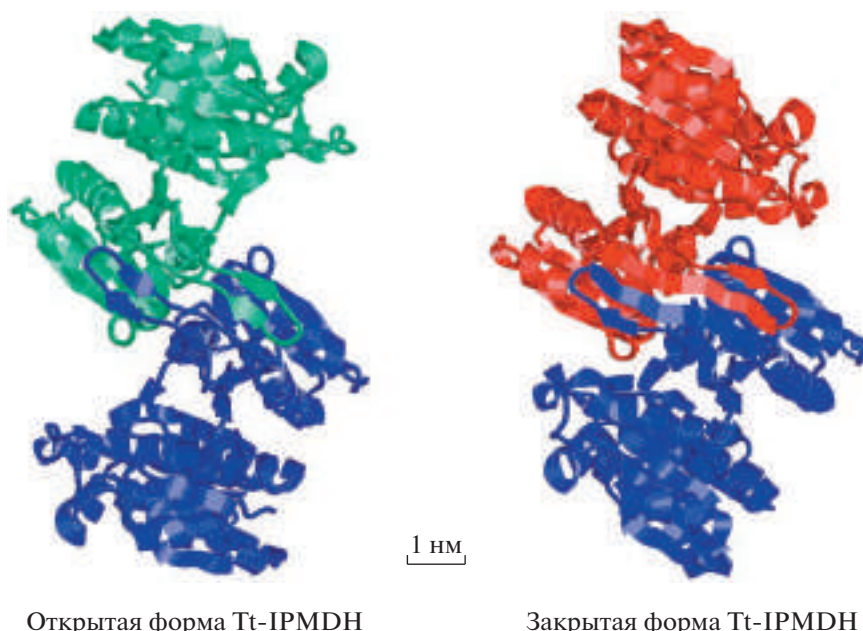
ВВЕДЕНИЕ

Малоугловое рентгеновское рассеяние (МУРР) является универсальным структурным методом исследования биологических макромолекул в растворе [1], позволяющим восстанавливать форму макромолекул в монодисперсных системах при разрешении $\sim 10\text{--}20$ Å. В последнее время в экспериментах МУРР широко применяется встроенная онлайн хроматографическая колонка (SEC-SAXS, SEC – Size Exclusion Chromatography, SAXS – Small-angle X-ray Scattering или МУРР), что помогает выделить сигнал от индивидуальных компонентов в белковых смесях или переходных комплексах. Однако даже при использовании SEC-SAXS-подхода не всегда удается полностью разделить компоненты и такие данные трудно интерпретировать.

Существуют разные подходы к решению задачи восстановления профилей рассеяния отдельных компонентов по данным МУРР, в частности это использование хемометрических методов [2] и эволюционного факторного анализа (EFA – Evolving Factor Analysis или ЭФА) [3, 4]. На основе последнего алгоритма разработали программу EFAMIX [5], которая позволяет в автоматическом режиме обрабатывать наборы SEC-SAXS данных

от белковых смесей и восстанавливать профили рассеяния от компонентов. Эффективность работы алгоритма была проверена на ряде теоретических и экспериментальных примеров в зависимости от количества компонентов, уровня шума в данных, степени асимметрии и соотношения амплитуд пиков концентрационных профилей компонентов для случая олигомерной белковой смеси. В частности, было показано, что по мере увеличения числа компонентов смеси увеличивается пороговое значение шума, выше которого восстановление индивидуальных профилей становится затруднительным.

Поскольку эффективность работы алгоритма ЭФА имеет определенные ограничения, зависящие от степени перекрывания концентрационных профилей компонентов, то в ряде случаев для определения вариантов “окон присутствия” компонентов и проверки результатов восстановления профилей рассеяния компонентов более удобно использовать интерактивный режим моделирования SEC-SAXS-данных. Это можно сделать с помощью программы CHROMIXS [6], ее графический интерфейс содержит специальное меню, задающее конфигурацию вызова для программы EFAMIX [5, 7].



Открытая форма Tt-IPMDH

Закрытая форма Tt-IPMDH

Рис. 1. Кристаллографические модели фермента Tt-IPMDH дикого типа (PDB ID: 2Y3Z) и его комплекса с субстратами Mn^{2+} , IPM и NADH (PDB ID: 4F7I). Ось симметрии второго порядка направлена перпендикулярно плоскости рисунка. Мономерные части структуры гомодимера выделены разными цветами.

Отметим, что обычно при использовании хроматографической колонки удастся разделять компоненты, молекулярные массы которых различаются в 2 и более раз. Например, это может быть олигомерная белковая смесь, содержащая мономеры, димеры и олигомеры более высоких порядков. В этом случае профили рассеяния от компонентов смеси сильно отличаются друг от друга (как по скорости спада кривых интенсивности рассеяния, так и по положению максимумов/минимумов на них). Но возникает вопрос, насколько эффективно будет работать алгоритм ЭФА, если кривые рассеяния достаточно близки друг к другу и соответствуют двум конформационным состояниям одного и того же белка.

В настоящей работе исследована такая белковая смесь с равными молекулярными массами компонентов на примере открытой (неактивной) и закрытой (активной) форм фермента 3-изопропилмалатдегидрогеназы (IPMDH – 3-isopropylmalate dehydrogenase) из *Thermus thermophilus* (Tt) [8].

СТРУКТУРА ФЕРМЕНТА 3-ИЗОПРОПИЛМАЛАТДЕГИДРОГЕНАЗЫ

Tt-IPMDH относится к классу β -гидроксикислотных окислительных декарбоксилаз и катализирует окисление и декарбоксилирование (2R,3S)-3-изопропилмалата (IPM) до 2-оксо-4-метилпентаноата в присутствии NAD и двухвалентного катиона (Mn^{2+} или Mg^{2+}) в биосинтетическом пути лейцина бактерий. Ранее были получены кри-

сталлографические модели Tt-IPMDH дикого типа (PDB ID: 2Y3Z) (открытая форма) и его комплекса с субстратами Mn^{2+} , IPM и NADH (PDB ID: 4F7I) (закрытая форма) [9, 10] (рис. 1). Они представляют собой димеры с общей молекулярной массой ~74 кДа, при этом каждый мономер состоит из двух примерно равных по размеру доменов. Данные модели были использованы для моделирования набора данных SEC-SAXS. Теоретические кривые рассеяния от моделей, рассчитанные с помощью программы CRY SOL [11], представлены на рис. 2. Кривая от открытой фор-

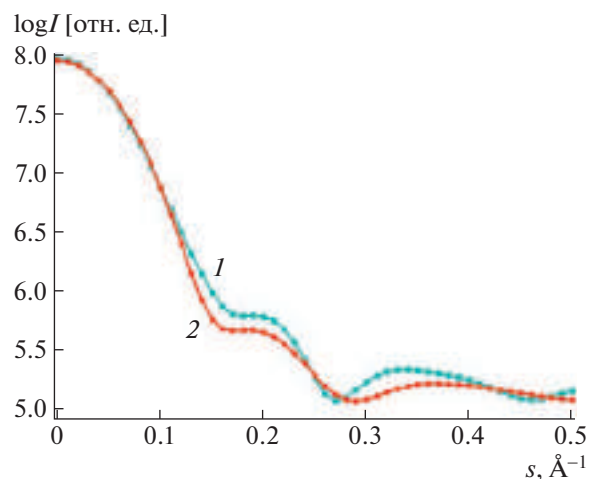


Рис. 2. Теоретические кривые рассеяния, рассчитанные от открытой (1) и закрытой (2) форм фермента Tt-IPMDH с помощью программы CRY SOL.

мы фермента Tt-IPMDH содержит более ярко выраженное “плечо” в угловом диапазоне $s = 0.15\text{--}0.25 \text{ \AA}^{-1}$ по сравнению с закрытой формой. В то же время радиусы инерции обоих структур близки и отличаются друг от друга всего на $1.5 \text{ \AA}$ ($R_g = 26.7 \text{ \AA}$ для закрытой формы и $R_g = 28.2 \text{ \AA}$ для открытой структуры). Ранее эксперименты МУРР подтвердили частичное закрытие доменов фермента Tt-IPMDH в присутствии ионов Mn^{2+} и Mg^{2+} , причем полученные данные хорошо приближались смесью активной и неактивной форм фермента [12].

МЕТОД ЭВОЛЮЦИОННОГО ФАКТОРНОГО АНАЛИЗА ДЛЯ АНАЛИЗА НАБОРА ДАННЫХ МУРР, ПОЛУЧЕННЫХ С ИСПОЛЬЗОВАНИЕМ ХРОМАТОГРАФИЧЕСКОЙ КОЛОНКИ

Эволюционный факторный анализ — это модельно-независимый подход для анализа матриц данных, составленных из одномерных кривых (в рассматриваемом случае данных SEC-SAXS) от многокомпонентных систем, в которых наблюдается последовательное, но неполное временное разделение компонентов.

Набор данных SEC-SAXS, состоящий из нескольких профилей рассеяния, снятых в разные моменты времени, описывается матрицей $\mathbf{A} = \{A_{ik}\} = \{I^{(k)}(s_i)\}$ ($i = 1, \dots, N$, $k = 1, \dots, K$, где N — число экспериментальных точек, K — число кривых МУРР, например общее число временных кадров, $s = 4\pi \sin \theta / \lambda$ — модуль вектора рассеяния, 2θ — угол рассеяния, λ — длина волны рентгеновского излучения. С помощью сингулярного разложения (SVD — Singular Value Decomposition) эта матрица может быть представлена как $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, где матрица $\mathbf{S}$ диагональная, а столбцы ортогональных матриц $\mathbf{U}$ и $\mathbf{V}$ — собственные векторы матриц $\mathbf{A}\mathbf{A}^T$ и $\mathbf{A}^T\mathbf{A}$ соответственно [13]. Основная идея алгоритма ЭФА заключается в том, чтобы проследить за рангом матрицы данных $\mathbf{A}$ в зависимости от количества принятых во внимание временных кадров. Для этого проводят SVD-разложение набора данных в прямом и обратном (временных) направлениях с последовательным увеличением размера матрицы данных. Далее по графикам зависимости сингулярных значений матрицы $\mathbf{S}$ от числа включенных в матрицу кривых рассеяния оценивают временные “окна присутствия” компонентов, когда соответствующие сингулярные значения начинают резко расти выше некоторой базовой линии или резко снижаться, приближаясь к базовой линии.

Примем во внимание, что матрица данных имеет вид $\mathbf{A} = \mathbf{I}\mathbf{C}$, где столбцы матрицы $\mathbf{I}$ представляют собой профили рассеяния компонентов, а матрица концентрации $\mathbf{C}$ выражается как

$\mathbf{C} = (\mathbf{U}^T\mathbf{I})^{-1}\mathbf{S}\mathbf{V}^T = \mathbf{R}\mathbf{V}^T$, где $\mathbf{R}$ — так называемая матрица вращения размером $M \times M$, где M — число компонентов в исследуемой системе. Последняя матрица неизвестна, но может быть найдена с помощью информации об окнах присутствия компонентов, полученной на предыдущем этапе. Взяв только те элементы сингулярных векторов $\mathbf{V}^T$ и матрицы $\mathbf{C}$, которые находятся вне диапазона концентрационного окна m -го компонента, можно последовательно восстановить один за другим столбцы матрицы $\mathbf{R}$, а затем вычислить элементы матрицы $\mathbf{C}$ [3]. Последним шагом алгоритма ЭФА является расчет матрицы $\mathbf{I}$, который проводится с использованием псевдо-обратной матрицы Мура–Пенроуза $\mathbf{C}^{-1}$, умноженной на матрицу данных $\mathbf{A}$.

Описанный выше неитеративный метод нахождения матрицы вращения не имеет ограничений на получаемые значения матрицы концентраций $\mathbf{C}$ и в некоторых случаях может приводить к отрицательным значениям концентраций, когда необходимо перепроверять автоматическую оценку “окон присутствия” компонентов, сканируя различные временные области и сравнивая полученные результаты. Однако альтернативные итеративные методы оценки матрицы $\mathbf{R}$, накладывающие ограничения на положительность матрицы $\mathbf{C}$ в процессе поиска, не всегда сходятся и имеют риск “застрять” в локальных минимумах, далеких от истинного решения.

РАЗДЕЛЕНИЕ ПЕРЕКРЫВАЮЩИХСЯ ВКЛАДОВ В ИНТЕНСИВНОСТЬ МУРР ОТ ОТКРЫТОЙ И ЗАКРЫТОЙ ФОРМ ФЕРМЕНТА Tt-IPMDH С ПОМОЩЬЮ АЛГОРИТМА ЭФА

Взяв известные структуры фермента Tt-IPMDH, рассчитали набор данных SEC-SAXS в виде линейной комбинации их теоретических кривых рассеяния с большим временным перекрытием пиков выхода компонентов (рис 3а). Поскольку любые реальные экспериментальные данные содержат шум, то необходимо было соответствующим образом добавить его к теоретическим наборам данных SEC-SAXS. Современные пиксельные 2D-детекторы (например, Pilatus, Eiger и др.) считаются счетчиками одиночных фотонов, и регистрируемые сигналы подчиняются статистике Пуассона. При этом во время процедуры моделирования шума проводить вычитание буфера не требуется, так как сумма двух случайных величин с пуассоновским распределением также будет иметь пуассоновское распределение. Однако необходимо учесть распространение ошибок при азимутальном усреднении двумерных данных, для чего использовали алгоритм генерации пуассоновского псевдослучайного шума, описанный в [14]. Расчет распространения ошибок при ази-

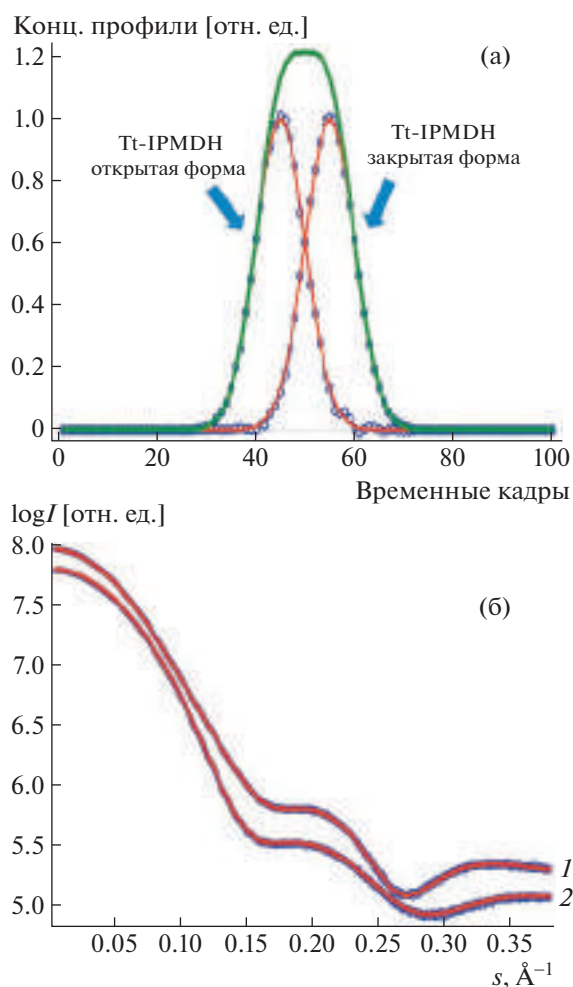


Рис. 3. Восстановление вкладов компонентов в интенсивность рассеяния по модельному набору данных SEC-SAXS с добавлением пуассоновского шума на уровне 5% с помощью алгоритма ЭФА. Концентрационные профили компонентов открытой и закрытой форм фермента Tt-IPMDH, теоретические (заданные) профили показаны сплошными линиями, профили, восстановленные с помощью программы EFAMIX, — кружками. Суммарный (оггибающий) концентрационный профиль показан сплошной линией (а). Теоретические (сплошные линии) и восстановленные программой EFAMIX (кружки) профили рассеяния открытой и закрытой форм фермента Tt-IPMDH. Профили рассеяния от открытой (1) и закрытой (2) форм фермента Tt-IPMDH смещены по вертикали для лучшей визуализации (б).

мутальном усреднении проводился для 2D-детектора с числом пикселей 2500×2500 . Для получения одномерных кривых рассеяния с различным отношением сигнал/шум исходные теоретические кривые масштабировали на коэффициент, соответствующий стандартным отклонениям пуассоновского шума в 1, 3, 5 и 10% в области максимальной интенсивности (начальная часть одномерной кривой). Рассматривая такие кривые как математические ожидания интенсивности,

для каждого пикселя 2D-детектора были рассчитаны псевдослучайные реализации пуассоновского шума. После этого значения интенсивности были усреднены в азимутальном направлении. Рассмотренные в работе уровни шума можно классифицировать как низкий шум (1%, что соответствует 10^4 фотонам в области, близкой к области прямого пучка на детекторе), умеренный (3 и 5% соответствуют 10^3 и 4×10^2 фотонам) и высокий (10% для 10^2 фотонов).

Для полученных наборов данных SEC-SAXS удалось с высокой точностью восстановить профили рассеяния двух конформационных состояний фермента с помощью программы EFAMIX в случае низкого и умеренного уровня шума (в пределах 1–5%), что подтверждает высокий потенциал метода ЭФА для разделения перекрывающихся вкладов компонентов в реальных наборах данных SEC-SAXS (рис. 36). При высоком уровне шума (10%) кривые рассеяния от открытой и закрытой форм фермента становятся уже практически неразличимыми, и в этом случае метод ЭФА не позволяет провести корректное разделение вкладов компонентов в интенсивность рассеяния.

Отметим, что в настоящее время использование хроматографических колонок, установленных непосредственно перед кюветой, входит в практику типичного оснащения синхротронных и даже лабораторных малоугловых станций. Разделение фракций с одинаковой молекулярной массой можно осуществлять за счет введения солевого градиента в буферном растворе, используя метод ионно-обменной хроматографии [5].

ЗАКЛЮЧЕНИЕ

Проанализированы модельные данные SEC-SAXS от белковой смеси двух конформационных состояний фермента Tt-IPMDH. Показано, что при наличии умеренного шума в данных на уровне 3–5% можно осуществить разделение вкладов в интенсивность малоуглового рассеяния отдельных компонентов смеси даже при наличии существенного перекрытия концентрационных профилей выхода компонентов из хроматографической колонки. Полученный результат демонстрирует высокую эффективность метода ЭФА при анализе данных SEC-SAXS от многокомпонентных систем.

Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках Федеральной научно-технической программы развития синхротронных и нейтронных исследований и исследовательской инфраструктуры на 2019–2027 годы (соглашение № 075-15-2021-1355 от 12 октября 2021 г.).

СПИСОК ЛИТЕРАТУРЫ

1. *Svergun D.I., Koch M.H.J., Timmins P.A., May R.P.* Small angle X-ray and neutron scattering from solutions of biological macromolecules. Oxford University Press, 2013. 358 p.
2. *Herranz-Trillo F., Groenning M., van Maarschalkerweerd A. et al.* // *Structure*. 2017. V. 25. P. 5. <https://doi.org/10.1016/j.str.2016.10.013>
3. *Keller H.R., Massart D.L.* // *Chemom. Intell. Lab. Syst.* 1992. V. 12. P. 209. [https://doi.org/10.1016/0169-7439\(92\)80002-L](https://doi.org/10.1016/0169-7439(92)80002-L)
4. *Hopkins J.B., Gillilan R.E., Skou S.J.* // *J. Appl. Cryst.* 2017. V. 50. P. 1545. <https://doi.org/10.1107/S1600576717011438>
5. *Konarev P.V., Graewert M.A., Jeffries C.Y. et al.* // *Protein Sci.* 2022. V. 31. P. 269. <https://doi.org/10.1002/pro.4237>
6. *Panjikovich A., Svergun D.I.* // *Bioinformatics*. 2018. V. 34. P. 1944. <https://doi.org/10.1093/bioinformatics/btx846>
7. *Konarev P.V., Volkov V.V.* // *Physics of Atomic Nuclei*. 2022. V. 85. P. 2127. <https://doi.org/10.1134/S1063778822090198>
8. *Hayashi-Iwasaki Y., Oshima T.* // *Methods Enzymol.* 2000. V. 324. P. 301. [https://doi.org/10.1016/s0076-6879\(00\)24240-7](https://doi.org/10.1016/s0076-6879(00)24240-7)
9. *Graczer E., Merlin A., Singh R.K. et al.* // *Mol. Biosyst.* 2011. V. 7. P. 1646. <https://doi.org/10.1039/C0MB00346H>
10. *Pallo A., Olah J., Graczer E. et al.* // *FEBS J.* 2014. V. 281. P. 5063. <https://doi.org/10.1111/febs.13044>
11. *Svergun D.I., Barberato C., Koch M.H.J.* // *J. Appl. Cryst.* 1995. V. 28. P. 768. <https://doi.org/10.1107/S0021889895007047>
12. *Graczer E., Konarev P.V., Szimler. T. et al.* // *FEBS Lett.* 2011. V. 585. P. 3297. <https://doi.org/10.1016/j.febslet.2011.09.013>
13. *Golub G.H., Reinsch C.* // *Numer. Math.* 1970. V. 14. P. 403. <https://doi.org/10.1007/bf02163027>
14. *Ahrens J.H., Dieter U.* // *ACM Trans Math Software*. 1982. V. 8. P. 163. <https://doi.org/10.1145/355993.355997>